AFA

Attorney Docket: YOR920000807US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): Brown et al.
Docket No.: YOR920000807US1
Serial No.: 09/713,075
Filing Date: November 15, 2000
Group: 2654
Examiner: Lamont M. Spooner

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Signature: Susan Fortuna Date: December 2, 2005

Title: System and Method for Finding the Most Likely Answer to a Natural Language Question

---

TRANSMITTAL OF APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith are the following documents relating to the above-identified patent application:

1. Appeal Brief; and
2. Copy of Notice of Appeal, filed on November 9, 2005, with copy of stamped return postcard indicating receipt of Notice by PTO on November 14, 2005.

There is an additional fee of $500 due in conjunction with this submission under 37 CFR §1.17(c). Please charge **IBM Corporation's Deposit Account No. 50-0510** the amount of $500 to cover this fee. In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **IBM Corporation's Deposit Account No. 50-0510** as required to correct the error. A duplicate copy of this letter is enclosed.

Respectfully,

Kevin M. Mason

Date: December 2, 2005

Kevin M. Mason
Attorney for Applicant(s)
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06824
(203) 255-6560

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s):	Brown et al.
Docket No.:	YOR920000807US1
Serial No.:	09/713,075
Filing Date:	November 15, 2000
Group:	2654
Examiner:	Lamont M. Spooner

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Signature: _____ Date: December 2, 2005

Title: System and Method for Finding the Most Likely Answer to a Natural Language Question

---

APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Applicants hereby appeal the final rejection dated September 9, 2005, of claims 1 through 12 of the above-identified patent application.

REAL PARTY IN INTEREST

An assignment was filed on November 4, 2005, assigning the present application to International Business Machines Corporation. The assignee, International Business Machines Corporation, is the real party in interest.

RELATED APPEALS AND INTERFERENCES

There are no related appeals and interferences.

STATUS OF CLAIMS

Claims 1 through 12 are pending in the above-identified patent application. Claims 1, 2, 7, 8, 10, and 11 remain rejected under 35 U.S.C. §102(b) as being anticipated by Braden-Harder et al. (United States Patent Number 5,933,822) and claims 3-6, 9, and 12 remain rejected under 35 U.S.C. §103(a) as being unpatentable over Braden-Harder et al. in view of Diamond (United States Patent Number 6,269,368).

STATUS OF AMENDMENTS

There have been no amendments filed subsequent to the final rejection.

SUMMARY OF CLAIMED SUBJECT MATTER

The present invention is directed to automated question answering that relates to the selection of an answer to a question from a pool of potential answers which are manually or automatically extracted from a large collection of textual documents (page 5, line 18, to page 7, line 17). The feature extraction component, a feature combination component, an answer selection component, and an answer presentation component, among others, are included (page 7, line 18, to page 9, line 6). The input to the system is a set of one or more natural language questions and a collection of textual documents (page 9, lines 15-22). The output is a (possibly ranked) set of factual answers to the questions, these answers being extracted from the document collection (page 9, line 23, to page 10, line 5).

STATEMENT OF GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1, 2, 7, 8, 10, and 11 are rejected under 35 U.S.C. §102(b) as being anticipated by Braden-Harder et al. and claims 3-6, 9, and 12 are rejected under 35 U.S.C. §103(a) as being unpatentable over Braden-Harder et al. in view of Diamond.

ARGUMENT

Independent Claims 1, 7 and 12

Independent claims 1 and 7 were rejected under 35 U.S.C. §102(b) as being anticipated by Braden-Harder et al. and claim 12 was rejected under 35 U.S.C. §103(a) as being unpatentable over Braden-Harder et al. in view of Diamond.

Regarding claims 1 and 7, the Examiner asserts that Braden-Harder discloses presenting the best scoring possible answer to the user with context from the passage containing the answer. Regarding claim 12, the Examiner asserts that Diamond teaches scoring each possible answer phrase, selecting one or more of the best scoring answer phrases, and displaying the answer phrases to the user. In the Response to Arguments section of the final Office Action, the Examiner asserts that the Applicants' argument that "Braden-Harder does not attempt to present answers with context from a relevant passage" is not claimed, and that the invention as claimed does not limit the "candidate list of passages of possible answers" to a phrase, a sentence, a paragraph, or a document. The Examiner further asserts that a "pool of potential answers...extracted from a large collection of textual documents" cannot be limited to a "snippet."

Applicants note that the present invention is directed to selecting *answers* to natural language questions "from a pool of potential answers which are manually or automatically extracted from a large collection of textual documents." (Page 1, lines 9-11.) An "answer" is defined as "a statement (either spoken or written) that is made in reply to a question or request or criticism or accusation." (See, dictionary.com) Thus, answers are ***not*** *documents* but are *short snippets or "snippets of text (e.g., phrases) which provide the exact answer to the question.*" (See, definition on page 1, lines 21-23, of the present disclosure.) Although Braden-Hader utilizes natural language processing to identify documents that are relevant to a query, Braden-Harder discloses that the processor "presents the retained ***documents*** to the user rank-ordered based on their score." (Col. 8, lines 2-4; emphasis added.) As with most conventional search engines, Braden-Harder simply returns documents. Braden-Harder does not attempt to present *answers, as defined in the present invention and as would be understood by a person of ordinary skill in the art, and does not attempt to present answers with context from a*

relevant passage. Independent claim 1 requires "presenting the best scoring possible *answer* to the user *with context from the passage* containing the answer." Thus, contrary to the Examiner's assertion, presenting answers with context from a relevant passage is claimed in the present disclosure. In addition, independent claim 7 requires an *answer selection module* and an *answer presentation module,* and independent claim 12 requires *scoring "each possible answer phrase, selecting one or more of the best scoring answer phrases, and displaying the answer phrases* to the user." Braden-Harder does not disclose or suggest any of these limitations.

Applicants also note that Diamond is directed to a system and method for improving information retrieval effectiveness by dynamically combining evidence information produced by a plurality of retrieval systems matching alternative representations of queries and documents. (See, Abstract.) As with most conventional search engines, Diamond simply returns documents. Diamond does not attempt to present *answers,* as defined in the present invention and as would be understood by a person of ordinary skill in the art, and does not attempt to present *answers with context from a relevant passage.*

Thus, Braden-Harder and Diamond, alone or in combination, do not disclose or suggest presenting the best scoring possible answer to the user with context from the passage containing the answer, as required by claim 1, do not disclose or suggest an answer selection module and an answer presentation module, as required by independent claim 7, and do not disclose or suggest scoring each possible answer phrase, selecting one or more of the best scoring answer phrases, and displaying the answer phrases to the user, as required by independent claim 12.

Claims 5 and 9

Claims 5 and 9 are rejected under 35 U.S.C. §103(a) as being unpatentable over Braden-Harder et al. in view of Diamond. In particular, the Examiner asserts that Braden-Harder discloses that "the parameters of the scoring function are learned by a machine learning algorithm."

Applicants note that the text cited by the Examiner is directed to "the manner through which our invention (Braden-Harder) compares and weights matching

logical form triples, and ranks corresponding documents." (Col. 17, lines 5-7.) Applicants could find no disclosure or suggestion by Braden-Harder, however, that "the ***parameters*** *of the scoring function* are learned by a machine learning algorithm."

Thus, Braden-Harder and Diamond, alone or in combination, do not disclose or suggest the parameters of the scoring function are learned by a machine learning algorithm, as required by claims 5 and 9.

Conclusion

The rejections of the cited claims under sections 102 and 103 in view of Braden-Harder and Diamond, alone or in combination, are therefore believed to be improper and should be withdrawn. The remaining rejected dependent claims are believed allowable for at least the reasons identified above with respect to the independent claims.

The attention of the Examiner and the Appeal Board to this matter is appreciated.

Respectfully,

Date: December 2, 2005

Kevin M. Mason
Attorney for Applicant(s)
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06824
(203) 255-6560

APPENDIX

1. A method for selecting answers to natural language questions from a collection of textual documents comprising the steps of:

extracting scoring features from a candidate list of passages of possible answers;

scoring the possible answers using the extracted features and a feature scoring function; and

presenting the best scoring possible answer to the user with context from the passage containing the answer.

2. A method as in claim 1, wherein the features used to score possible answers consists of one or more of the following features: a semantic type of a current suspected answer, a position of the suspected answer among all suspected answers within all document passages, a position of the suspected answer among all suspected answers within the given passage, a number of suspected answers of a given semantic type retrieved within a given passage, a number of words in a suspected answer that do not appear in the user question, a position of the semantic type in the list of potential semantic types for the question, an average distance in words between the beginning of the potential answer and the words in the question that also appear in the passage, a passage relevance as computed by the information retrieval engine, a frequency of a given potential answer on the list, a semantic relation between words from the question and words from the potential answer, and a strength score that is a function of the relevance score.

3. A method as in claim 2, wherein the feature scoring function is a linear combination of weighted features.

4. A method as in claim 3, wherein the parameters of the scoring function are manually determined.

5. A method as in claim 3, wherein the parameters of the scoring function are learned by a machine learning algorithm.

6. A method as in claim 1 where the candidate list of passages of possible answers is obtained from the collection of documents using an information retrieval engine

7. A computer system that extracts answers to natural language questions from a large collection of textual documents consisting of one or more of the following modules:

a feature extraction module;

a feature combination module, containing a "feature extraction" and "compute composite score" components;

an answer selection module; and

an answer presentation module.

8. A computer system, as in claim 7, wherein the feature extraction module extracts one or more of the following features: a semantic type of the current suspected answer, a position of the suspected answer among all suspected answers within all document passages, a position of the suspected answer among all suspected answers within the given passage, a number of suspected answers of a given semantic type retrieved within a given passage, a number of words in an suspected answer that do not appear in the user question, a position of the semantic type in the list of potential semantic types for the question, an average distance in words between the beginning of the potential answer and the words in the question that also appear in the passage, a passage relevance as computed the retrieval engine, a frequency of a given potential answer on the list, a semantic relation between words from the question and words from the potential answer, and a strength score that is a function of the relevance score.

9. A computer system as in claim 7, wherein the feature combination module employs a feature scoring function with parameters learned by a machine learning method.

10. A computer system as in claim 7, wherein the answer selection module selects the answer with the best score obtained from the feature combination module.

11. A computer system as in claim 7, wherein the answer presentation module shows the top scored answer within the context as specified by a user or a system.

12. A computer program product that performs the steps of:

determining a feature scoring function during a training phase either manually or via a machine learning algorithm applied to a set of training questions, corresponding answer passages, and certain extracted features; and

during an execution phase, extracting certain features from questions and corresponding possible answer phrases, applying the feature scoring function determined during the training phase to score each possible answers phrase, selecting one or more of the best scoring answer phrases, and displaying the answer phrases to the user with optional additional context from the answer passages.

EVIDENCE APPENDIX

There is no evidence submitted pursuant to § 1.130, 1.131, or 1.132 or entered by the Examiner and relied upon by appellant.

RELATED PROCEEDINGS APPENDIX

There are no known decisions rendered by a court or the Board in any proceeding identified pursuant to paragraph (c)(1)(ii) of 37 CFR 41.37.